

Mapping from UMLS CUI into the OHDSI Vocabulary Release 1.0

Juan M. Banda

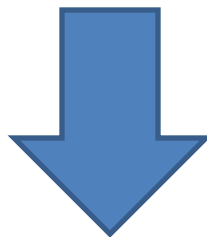




What are we trying to do?



Concept Unique Identifiers (CUI)



OHDSI concept_id's



ATHENA Standardized Vocabularies



Challenges

- UMLS (2017AB) has 131 English vocabularies
- OHDSI Vocabulary (Jan-2018) has 71+ vocabularies
- Some versioning issues



The Real Challenge

- OHDIS concept_code which is the source code (in RxNorm, SNOMED) can be overlapping between different vocabularies
- Filtering on a per-vocabulary basis is needed



Our approach

- 1) Identify matching vocabularies on both sides (HPCS, ICD9, ICD10, SNOMED are easy)

UMLS2017AB – MRSAB table

RSAB	SON	SVER	SLC	SSN	SCIT
HCPT	HCPCS Version of Current Procedural Terminology (CPT),...	2017	Cynthia Hake;CMS HCPCS Workgroup Chair;Centers for ...	CPT in HCPCS	;;;Version of Physicians' Current Procedural Terminology...
HGNC	HUGO Gene Nomenclature Committee, 2017_05	2017_05	Elsbeth Bruford, PhD;Group Co-ordinator;HUGO Gene No...	HUGO Gene Nomenclature Committee	Gray KA, Yates B, Seal RL, Wright MW, Bruford EA;;HGN...
HL7V2.5	HL7 Vocabulary Version 2.5, 2003_08_30	2003_08_30	Health Level Seven;;;3300 Washtenaw Avenue;;Suite 22...	HL7 Version 2.5	Mark McDougall, Executive Director, Health Level Seven; ...
HL7V3.0	HL7 Vocabulary Version 3.0, 2016_07	2016_07	Health Level Seven International;;;3300 Washtenaw Ave...	HL7 Version 3.0	;;;Health Level Seven (HL7) Vocabulary;;;Ann Arbor, M...
HPO	Human Phenotype Ontology, 2017_06_28	2017_06_28	Dr. Peter N. Robinson;Professor of Computational Biolog...	HPO	;;;The Human Phenotype Ontology (HPO);;;The Huma...
ICD10	ICD10, 1998	1998	;;Office of Publications, World Health Organization;1211 ...	ICD-10	;;;International Statistical Classification of Diseases and ...
ICD10AE	ICD10, American English Equivalents, 1998	1998	;;Office of Publications, World Health Organization;1211 ...	ICD-10 Am Engl	;;;International Statistical Classification of Diseases and ...
ICD10AM	International Statistical Classification of Diseases and Rel...	2000	;;National Centre for Classification in Health, University o...	ICD-10 Austral Mod	;;National Centre for Classification in Health, University o...
ICD10AMAE	International Statistical Classification of Diseases and Rel...	2000	;;National Centre for Classification in Health University of...	ICD-10 Austral Mod Am Engl	;;;International Statistical Classification of Diseases and ...
ICD10CM	International Classification of Diseases, 10th Edition, Cini...	2018	Donna Pickett;Medical Classification Administrator;Nation...	ICD-10-CM	;;National Center for Health Statistics (NCHS), under aut...
ICD10PCS	ICD-10-PCS, 2018	2018	Pat Brooks;Senior Technical Advisor;Center for Medicare ...	ICD-10-PCS	;;Centers for Medicare and Medicaid Services;;Internatio...
ICD9CM	International Classification of Diseases, Ninth Revision, C...	2014	Patricia Brooks;Contact for Procedures;Health Care Fina...	ICD-9-CM	;;National Center for Health Statistics (NCHS);;ICD-9-CM...
ICD9CM	International Classification of Diseases, Ninth Revision, C...	2000_10_10	Patricia Brooks;Contact for Procedures;Health Care Fina...	ICD-9-CM	;;National Center for Health Statistics (NCHS);;ICD-9-CM...

OHDSI Vocabulary – Vocabulary table

HCPCS	Healthcare Common Procedure Coding System (CMS)	2017 Alpha Numeric HCPCS File
HES Specialty	Hospital Episode Statistics Specialty (NHS)	Not implemented yet
ICD10	International Classification of Diseases, Tenth Revision (...)	2016 Release
ICD10CM	International Classification of Diseases, Tenth Revision, ...	ICD10CM FY2018 code descriptions
ICD10PCS	ICD-10 Procedure Coding System (CMS)	ICD10PCS 20160518
ICD9CM	International Classification of Diseases, Ninth Revision, C...	ICD9CM v32 master descriptions
ICD9Proc	International Classification of Diseases, Ninth Revision, C...	ICD9CM v32 master descriptions
ICD03	ICD-O-3	ICDO3 SEER Site/Histology Released 09/18/2015
ISBT	ISBT 128 Product Description Code Database	7.9.0
ISBT Attribute	ISBT 128 Product Description Code Database	7.9.0
LOINC	Logical Observation Identifiers Names and Codes (Regen...	LOINC 2.61
MDC	Major Diagnostic Categories (CMS)	2013-01-06
Meas Type	OMOP Measurement Type	
MedDRA	Medical Dictionary for Regulatory Activities (MSSO)	MedDRA version 19.1
MeSH	Medical Subject Headings (NLM)	2016 Release



Our approach (2)

2) Obtain CUI via concept_code (OHDSI) and code (UMLS)

CUI	concept_id	vocabulary_id
C0000039	1592753	RxNorm
C0000266	19041783	RxNorm
C0000294	1354698	RxNorm
C0000378	44814542	RxNorm
C0000392	19123344	RxNorm
C0000464	19016670	RxNorm
C0000473	19018384	RxNorm

UMLS via CUI

CUI	LAT	TS	LUI	STT	SUI	ISPREF	AUI	SAUI	SCUI	SDUI	SAB	TTY	CODE	STR	SRL	SUPPRESS	CVF
C0000039	ENG	P	L0000039	PF	S17175117	N	A28315139	9194921	1926948	(NULL)	RXNORM	IN	1926948	1,2-dipalmitoylphosphatidylcholine	0	N	(NULL)

OHDSI Vocab via concept_id

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason
1592753	1,2-dipalmitoylphosphatidylcholine	Drug	RxNorm	Ingredient	S	1926948	2017-08-07	2099-12-31	



Version 1 release

- First release of mappings include:

- **10 Standard vocabularies**

- ICD9
- ICD10
- HCPCS
- SNOMED
- RxNorm
- CPT
- MedDRA
- NDFRT
- LOINC



Results of V1 mapping

- Total number of concepts mapped: 1,119,625

vocabulary_id	number_of_mapped_concepts
CPT4	39,947
HCPCS	6,376
ICD10CM	101,725
ICD10PCS	180,450
ICD9CM	16,376
LOINC	124,813
MedDRA	52,953
NDFRT	23,345
RxNorm	202,627
SNOMED	371,013



Results of V1 mappings

- Actual number of Source vocabulary concepts in the OHDSI vocabulary

vocabulary_id	number_of_original_concepts
CPT4	15,446
HCPCS	8,129
ICD10CM	109,389
ICD10PCS	182,106
ICD9CM	18,672
LOINC	135,899
MedDRA	99,645
NDFRT	37,411
RxNorm	278,569
SNOMED	801,303



Results of V1 mappings

- Deltas

Vocabulary	Difference
CPT4	24,501
HCPCS	(1,753)
ICD10CM	(7,664)
ICD10PCS	(1,656)
ICD9CM	(2,296)
LOINC	(11,086)
MedDRA	(46,692)
NDFRT	(14,066)
RxNorm	(75,942)
SNOMED	(430,290)



Why does the CPT mapping explode?

CPT only has 13,265 unique `concept_id`'s (found in the OHDSI concept table) but these map to 39,947 cuis - Why?

CUI	SAB	CODE	STR	🔑 OHDSI_concept_id	OHDSI_concept_name	🌿 OHDSI_concept_code
C0373341	CPT	76825	Echocardiography, fetal, cardiovascular system, real time with image documentation (2D), with or without M-mode recording	2722250	Echocardiography, fetal, cardiovascular system, real tim...	76825
C3525406	CPT	76825	Real time echocardiography of fetal cardiovascular system	2722250	Echocardiography, fetal, cardiovascular system, real tim...	76825
C3703899	CPT	76825	Ultrasound of fetal heart blood flow	2722250	Echocardiography, fetal, cardiovascular system, real tim...	76825
C3866122	CPT	76825	Real time echocardiography of fetal cardiovascular system with 2D image documentation, without M-mode recording	2722250	Echocardiography, fetal, cardiovascular system, real tim...	76825
C3866123	CPT	76825	Real time echocardiography of fetal cardiovascular system with 2D image documentation and M-mode recording	2722250	Echocardiography, fetal, cardiovascular system, real tim...	76825



Why HCPCS has entries missing?

- OHDSI has 8,129 unique concept_codes
- UMLS only has 6,508 unique CODES

Result #1 (28x1,631)															
concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason	CUI	LAT	TS	LUI	STT	SUI
2614678	AMBULANCE TRANSPORT PROVIDED BETWEEN THE HOU...	Observation	HCPCS	HCPCS		A0800	1970-01-01	2011-01-09	D	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
2614775	MALE EXTERNAL CATHETER WITH INTEGRAL COLLECTIO...	Device	HCPCS	HCPCS		A4348	1970-01-01	2011-01-09	D	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
2614785	URINARY SUSPENSORY WITHOUT LEG BAG	Device	HCPCS	HCPCS		A4359	1970-01-01	2011-01-09	D	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
2614791	ADHESIVE REMOVER WIPES, ANY TYPE, PER 50	Device	HCPCS	HCPCS		A4365	1970-01-01	2014-01-01	D	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
2614866	ABDOMINAL DRESSING HOLDER	Device	HCPCS	HCPCS		A4462	1970-01-01	2011-01-09	D	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
2614942	REPLACEMENT BATTERY FOR EXTERNAL INFUSION PUMP	Device	HCPCS	HCPCS		A4632	1970-01-01	2011-01-09	D	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)	(NULL)



How about SNOMED?

- OHDSI has 801,303 unique concept_codes
- UMLS only has 449,474 unique CODES

Result #1 (28x358,421)

concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason	CUI	LAT	TS	LUI	STT
37392117	Urine tryptophan:creatinine ratio	Observation	SNOMED	Observable Entity	S	1000001000000103	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
46139744	Nu-Hope belt left small 6460C 70cm-78cm length, 225mm...	Drug	SNOMED	Pharma/Biol Product		100000110000001109	1970-01-01	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
37398455	Urine threonine:creatinine ratio	Observation	SNOMED	Observable Entity	S	10000110000000101	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
46004596	Nu-Hope belt right large 6457A 90cm-100cm length, 200...	Drug	SNOMED	Pharma/Biol Product		100001110000001105	1970-01-01	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
37392118	Urine taurine:creatinine ratio	Observation	SNOMED	Observable Entity	S	10000210000000107	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
46004597	Nu-Hope belt left small 6460DC 70cm-78cm length, 225m...	Drug	SNOMED	Pharma/Biol Product		100002110000001104	1970-01-01	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
37392119	Urine phenylalanine:creatinine ratio	Observation	SNOMED	Observable Entity	S	10000310000000109	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
46139745	Nu-Hope belt left small 6460DC 70cm-78cm length, 225m...	Drug	SNOMED	Pharma/Biol Product		100003110000001107	1970-01-01	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
37392120	Urine ornithine:creatinine ratio	Observation	SNOMED	Observable Entity	S	10000410000000100	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
46139746	Nu-Hope belt right large 6457A 90cm-100cm length, 200...	Drug	SNOMED	Pharma/Biol Product		100004110000001100	1970-01-01	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
37392121	Urine methionine:creatinine ratio	Observation	SNOMED	Observable Entity	S	10000510000000102	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
46004598	Nu-Hope belt left small 6460E 70cm-78cm length, 225mm ...	Drug	SNOMED	Pharma/Biol Product		100005110000001101	1970-01-01	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)
37392122	Urine lysine:creatinine ratio	Observation	SNOMED	Observable Entity	S	10000610000000104	2016-07-31	2099-12-31		(NULL)	(NULL)	(NULL)	(NULL)	(NULL)



Actual concepts in UMLS

vocabulary_id	number_of_original_concepts	In UMLS
CPT4	15446	13786
HCPCS	8129	6508
ICD10CM	109389	94518
ICD10PCS	182106	190177
ICD9CM	18672	22406
LOINC	135899	205523
MedDRA	99645	79571
NDFRT	37411	45364
RxNorm	278569	206957
SNOMED	801303	449474



Actual deltas with UMLS

Vocabulary	Difference	Diff w UMLS
CPT4	24,501	26,161
HCPCS	(1,753)	(132)
ICD10CM	(7,664)	7,207
ICD10PCS	(1,656)	(9,727)
ICD9CM	(2,296)	(6,030)
LOINC	(11,086)	(80,710)
MedDRA	(46,692)	(26,618)
NDFRT	(14,066)	(22,019)
RxNorm	(75,942)	(4,330)
SNOMED	(430,290)	(78,461)



Is this really a problem?

- If your annotator uses UMLS CUIs

.... Probably not.... But



V1 availability

- Two flavors

- Mapping table

- <https://github.com/jmbanda/OHDSIconceptid2cui/blob/master/CUItoOHDSIv1.zip>

- Mapping Script

- https://github.com/jmbanda/OHDSIconceptid2cui/blob/master/cui2ohdsi_concept_idv1.0.sql



What is next ?

- Fully customizable R package
 - Currently only works for MySQL
 - Work needs to be done to cover OHDSI supported RDBMs
- Mapping update
 - Now R package uses a nifty little file, which users can add and share!



Question?

Thank you for your time!

jmbanda@stanford.edu